Equity and Excellence in Research Funding

Article (Published Version)

Hicks, Diana M and Katz, Sylvan (2011) Equity and Excellence in Research Funding. Minerva, 49 (2). pp. 137-151. ISSN 0026-4695

Minerva
DOI 10.1007/s11024-011-9170-6

Equity and Excellence in Research Funding

Diana Hicks · J. Sylvan Katz

Abstract The tension between equity and excellence is fundamental in science policy. This tension might appear to be resolved through the use of merit-based evaluation as a criterion for research funding. This is not the case. Merit-based decision making alone is insufficient because of inequality aversion, a fundamental tendency of people to avoid extremely unequal distributions. The distribution of performance in science is extremely unequal, and no decision maker with the power to establish a distribution of public money would dare to match the level of inequality in research performance. We argue that decision makers who increase concentration of resources because they accept that research resources should be distributed according to merit probably implement less inequality than would be justified by differences in research performance. Here we show that the consequences are likely to be suppression of incentives for the very best scientists. The consequences for the performance of a national research system may be substantial. Decision makers are unaware of the issue, as they operate with distributional assumptions of normality that guide our everyday intuitions.

Keywords Research excellence · Distributional equity · Science policy · Power law · Lotka

D. Hicks (✉)
School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA
e-mail: dhicks@gatech.edu

J. S. Katz
SPRU, The Freeman Centre, University of Sussex, Brighton, East Sussex BN1 9QE, UK

Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan,
Saskatoon, SK S7N 5E6, Canada

Published online: 28 April 2011

Introduction

The tension between excellence and equity in university research performance and funding is a core science policy concern explored in a vast literature of descriptive essays rich in contextual narrative. In contrast, this paper takes a somewhat formal, quantitative and abstract approach to the topic. We offer a different approach in the hope that our perspective can advance understanding of how to distribute research funding in order to maximize a nation's international-level research excellence. With many countries entering a period of austerity, greater understanding of this issue is urgently needed. In this paper, we cannot solve this problem, rather we simply lay some of the groundwork for a future research agenda.

We approach the issue from a systems perspective. To do this we consider the distribution of performance and funding within a research system. Distributional thinking is uncommon, but arguably quite useful in thinking about science policy problems. For example, sometimes research is described as "winner take all" because a few people are responsible for a very high percentage of the advances. Winners are thus highly visible. Nevertheless, data demonstrate that there is a *distribution* in performance, meaning that some people do perform at an intermediate level. Because the "winner take all" shorthand misses this, it may not be the best basis for policy. As a second example, powerful statistical techniques are available to analyze normally distributed data. Arguably, this leads to a subtle problem in which all data is seen as somehow normally distributed. Scientific performance and therefore much data in science and technology studies are distributed in a power law fashion. This data is invariably referred to as "highly skewed." Skew means asymmetrical. The unspoken and inaccurate subtext in the common usage is that the data are some sort of badly misbehaving normal distribution. Not true. The data are stranger and more difficult to handle than that because the mean is not a useful statistic and variance is often infinite. It is the differences between the characteristics of the normal and power law distribution and their possible consequences for science policy that this paper articulates.

The argument proceeds as follows. The paper begins by discussing the distinctions commonly made between probability distributions. There follows a review of the empirical evidence that a power law distribution characterizes research performance. Next, tensions between excellence and equity in the distribution of research resources are discussed. Such tensions likely create a mismatch between the distribution of research resources that would be ideal from the viewpoint of maximizing research excellence and actual distributions based on merit. We explore how this might reduce system performance.

The argument in the paper is fairly stylized in discussing the simple dyad of "funders" and "researchers" without committing to a particular unit of analysis. Units of analysis can be tricky when discussing research funding, especially on the researcher side. Universities are an obvious player, and research money is allocated to them by government bodies in centralized systems. However, individuals are also players, though the individual "PI" is something of a fiction in that a person's capacity for obtaining competitively awarded grant funding will be shaped by other possible units of analysis – the department that provides a more or less conducive environment for research; groups, that is colleagues in various stages of career development who collaborate with and work for the person, and in so doing shape their ideas and help to execute their research; and the university, whose managerial capacity is essential to qualify its employees as potential grant recipients. To whom then is funding awarded – PI, group, department or university?

For better or worse, data availability often shapes choice of unit of analysis. Data at departmental level emerges from national evaluation systems which tend to reify this level. University level data is more broadly available internationally. But committing to one unit of analysis may preclude full understanding. Geuna's study of the economics of university research at the university level illustrated this. He was forced to abandon consistency in unit of analysis to explore micro-dynamics of funding at the level of research centers to understand the implications of grant mechanisms for the evolution of universities (Guena 1998, p.186). The discussion here is largely conceptual and cannot hope to resolve these difficulties; therefore a stylized funder/research dyad is used.

There is an important caveat to the argument of the paper. The argument applies best to publicly funded university research. Universities teach as well as perform research, and the distribution of performance in teaching quite likely differs from the distribution of performance in research. Therefore, we do not address the distribution of resources to universities as a whole, just the distribution of research resources.

The Characteristics of Probability Distributions

Probability distributions are embedded in analytical work and, we would argue, in our assumptions about how the world works. The most powerful statistical techniques assume data is in a normal distribution or a variant thereof. Analytical training embeds this view of the world within us, and analysts seek to transform any dataset into a normally distributed dataset so that their techniques can be used. Our everyday experience leads us to unconsciously assume most things are distributed like the things we know best, such as height or intelligence, in which a few people are short, a few are quite tall (but nobody is 20 feet tall) and most people are somewhere in the middle. Unfortunately, when the world does not conform to the normal distribution, our analytical and everyday assumptions can obscure understanding.

Probability distributions are commonly distinguished along several dimensions including mathematical descriptions, symmetry and tails. The fundamental description of a distribution is its mathematical function: normal, exponential, power law (in specific applications also referred to as Pareto, Lotka, Zipf), Poisson, negative binomial, chi-square etc. However, this, the most precise characterization of distributions, does not exhaust the useful distinctions that can be made. Symmetry is important in analytical work and is the basis for a second categorization. Asymmetric distributions are labeled "skew," or in the case of data in the realm of science and technology - "highly skewed". The very term embodies an assumption that the distribution in question is a misbehaving normal distribution which technique (e.g logarithmic transformation) can regularize.

A third increasingly common categorization addresses the outlying values in a distribution or the “tail.” When financial analysts found that disaster struck their portfolios more often than their models predicted, they traced the trouble to their distributional assumptions. Research in risk management now focuses on “fat tail” distributions in which the frequency of extreme values at the high or low end is higher than would be predicted by the normal or exponential distributions, which are both “thin tailed” distributions (they are related, the high end of a normal distribution decays exponentially). Work on fat tailed distributions is also propelled by research on computer networks which have fat tailed characteristics. Power law distributions have fat tails; normal distributions have thin tails.

The power law is a very unequal distribution. Figure 1 illustrates the difference in inequality between a normal and power law distribution as measured by the Gini coefficient.[1] Figure 1a shows a power law and a normal distribution of resources. The distributions were constructed so that they contain the same number of people, 988, and cover the same range: in both distributions somebody has $1 of resources and one person has $100 of resources. The normal distribution has a mean of 50 and a standard deviation of 20. The power law distribution has a constant of 231 and an exponent of 1.1[2]. The Lorenz curves for these distributions are displayed in Figure 1b. The Lorenz curve for the normal distribution is much closer to the line of equality than is the curve for the power law distribution. Therefore, the Gini coefficient of the normal distribution at 0.21 is smaller than the Gini coefficient of the power law distribution which is 0.66.[3] The Gini coefficient tells us that those seeking egalitarian distributions will prefer the normal distribution over the power law distribution.

Power Law Distribution of Research Performance

Alfred Lotka, an employee of the Metropolitan Life Insurance Company in New York, was the first to examine the frequency distribution of scientific productivity. In 1926, Lotka counted the number of papers published by individuals with names beginning with the letters A and B in 10 years of *Chemical Abstracts*, and the number of papers in an index of important physics papers (Lotka 1926). He

[1] The Gini coefficient is a measure of inequality that ranges between 0 and 1 with 0 representing perfect equality where everyone has the same resources and 1 representing perfect inequality where one person has all the income while everyone else has zero income. The Gini coefficient was originally based on the average of the absolute differences between all pairs of incomes in the distribution and defined to be the ratio of half of that average to the mean of the distribution (Dorfman, 1979). Today the Gini coefficient is commonly explained in geometric terms using Lorenz curves. For any income distribution, the Lorenz curve is constructed from data ordered ascending by share of resources. It plots the cumulative share of income against the cumulative share of population. If everyone had the same income, the curve would be a straight line at a 45 degree angle. This line is plotted on graphs of Lorenz curves and is called the “line of equality.” If one person had all the income, the curve would be a vertical line at 100% of the population. The Gini coefficient is a ratio between two areas: 1) the area between the line of equality and the Lorenz curve and 2) the total area under the line of equality.

[2] The equations for the power law is $p(x) = Cx^{-n}$ for $x > x_{min}$

[3] Gini coefficients calculated using Wessa, P. (2008), Free Statistics Software, Office for Research Development and Education, version 1.1.22-r4, URL http://www.wessa.net/

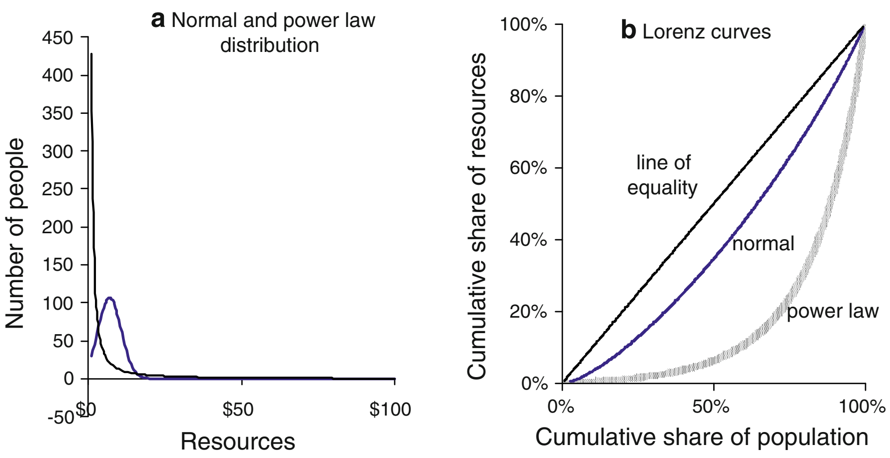


Fig. 1 Distributions and Lorenz curves for a normal and power law distribution

reported an "inverse square law of scientific productivity" according to which "the proportion of all contributors who contribute a single item should be just over 60 per cent." In other words, he found scientific productivity followed a power law.

Since then, a great deal of work has examined distributions of scientific productivity at all levels of analysis from individual to national and has calculated the values of the exponent on the power law. Recently questions have been raised about the methods used to empirically detect and characterize power laws, specifically Clauset et al. believe standard methods are not refined enough and too many distributions are being identified as power laws. According to Clauset et al. (2009), "standard methods such as least-squares fitting are known to produce systematically biased estimates of parameters for power-law distributions." Clauset et al. use maximum likelihood methods and the Kolmogorov-Smirnov statistic to reexamine twenty datasets previously found to follow a power law, including scientific productivity. They find moderate support for the identification of the distributions of author productivity and citations as a power law. We conclude that even using the most advanced methods, scientific productivity and citations are likely distributed in a power law.

In this case, the power law distribution reflects an underlying mechanism of cumulative advantage. The sociologist Robert Merton first identified this mechanism and called it the Matthew effect. More recently, the term "preferential attachment" is used, meaning that a quantity is distributed among individuals according to how much they already have. For example, well cited papers attract more citations and uncited papers continue to be ignored.

Equity versus Excellence in Research Funding Distribution

The role of cumulative advantage in scientific performance establishes a conflict between efficiency and equity considerations in the public funding of research. This conflict is endemic (Feller 2001). It was visible in the post World War II debate in the United States over the establishment of the National Science Foundation. Senator Harley Kilgore, from West Virginia, who worried about concentrations of power, held hearings on general government support of research and development and produced a report and a bill that proposed a well-funded agency that, among other things, would distribute research funding largely on a state-by-state formula. Vannevar Bush, an MIT engineer, president of the Carnegie Institution and leader of the wartime R&D effort, believed, among other things, that the notion of geographic distribution was inconsistent with maintaining a high quality of science (Morin 1993, 18-21). After much debate and delay, the National Science Foundation was established following the Bush vision, awarding grants based on peer review judgments of scientific excellence.

Several decades later Congress reacted badly to the extremely unequal distribution of Federal research resources across states (California and Massachusetts get far more in Federal research support than any other state). NSF was directed to avoid "undue concentration" of research and education and so established EPSCOR so that researchers in states at the bottom of the performance distribution could compete amongst themselves for a pool of research funding with the idea that this would build capacity and eventually states would "graduate" from the program (Wu 2010). Congressional earmarking of research funding often is justified as providing more distributional equity in research funding (Feller 2001).

More recently in Europe, Framework program funding has required that collaborative partners from weaker scientific countries be included on proposals with the aim of strengthening capacity. European research funding has also been subject to pressure for *juste retour* - or the principle that the distribution of European funding across countries should match the contributions countries make to the European budget. The European Research Council was established to escape this by quite explicitly funding research only on criteria of peer reviewed excellence irrespective of geographic considerations.

Beyond political and institutional self-interest, there are likely deep seated psychological preferences underpinning the high public value placed on equity. The results of behavioral economics and most recently neurobiology suggest that egalitarian preferences are quite deeply ingrained in human nature (Amiel & Cowell 1999; Dawes et al. 2007; Tricomi et al. 2010). The fundamental preference for greater equality in resource distribution within a group of which one is a part is called inequality aversion. Given our aversion to inequality, even decision makers seeking to foster research excellence through peer review processes would be extremely uncomfortable with the degree of inequality in resource allocation that would be required to match the performance distribution. The resulting mismatch between performance and resource distributions would compromise the efficiency of the allocation.

How do funders move resources towards a power law when they wish to more closely match the distribution of output? In many countries outside the United States, a considerable portion of core research resources are allocated to national universities by a decision maker in an annual funding round. In the past the funding was allocated using a formula based on student and/or faculty numbers. This method satisfies concerns about equity. The trend is to introduce systems in which part of the allocation is based on assessments of research output. The stated goals of governments introducing these systems include encouraging research excellence.

Another class of mechanism might be termed "select few" policies, meaning give large amounts of money to a few contenders. This is done in centers of excellence schemes (Japan and Germany currently attempting) and in large grants to young people (examples: NSF Career awards and European grants to young people). Injecting large amounts of money into a system in this way can in itself shift the overall shape of the distribution. In addition, there is a multiplier effect. The "select few" winners of one competition will carry the halo of being one of the select few into future funding competitions, enhancing their subsequent chances of success.

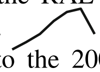
Competitive grant programs with low success rates are also a kind of "select few" policy. Geuna pointed out that a cumulative mechanism is also at work when multiple grant programs are available. Geuna found that winners of national grants felt they were more likely to get EU grants and vice versa (Geuna 1998). Cumulative mechanisms in growing systems are one of the known generators of fat tail, power law distributions. Such mechanisms have long been referred to as the "Matthew effect" in science studies.

Note that within this framework, innovation prizes can be viewed as a power law compatible distribution system. In particular, the X Prize Foundation "creates and manages prizes that drive innovators to solve some of the greatest challenges facing the world today." Their mission is "to bring about radical breakthroughs for the benefit of humanity" (xprize.org). They do this by establishing high-reward competitions to solve challenging scientific and engineering problems. Their first prize was the Ansari X Prize of $10 million which went to Burt Rutan for the first private suborbital space flight. In this process, what would constitute excellence must be pre-specified, and competitions are more innovation than science oriented. Because large amounts of money are given to the top performer, prizes have the potential to greatly increase inequality in resource distribution. Congress has recently allowed agencies to begin offering prizes so these two sentences are now out of date.

Two legitimate sets of public values contest for primacy in the allocation of research funding. Governments seek to maximize research excellence by distributing resources based on merit. Governments also enhance equity with a broad geographic distribution of research funds to support economic development, strengthen research and enhance diversity and participation in the research enterprise. Here we suggest that it is quite likely that our merit-based distributions are shaped by equity considerations, and that this might not be optimal.

Research Resource Distribution Example

In this section, we examine one process for distributing research resources, the RAE exercise in the UK, because it illustrates the tendency to avoid distributing resources in a power law. The British government periodically evaluates the research in its universities through the Research Assessment Exercise (RAE) in order to inform the distribution of general support for research to universities. In 2001, the RAE was a peer review evaluation of the research output of each department on a seven point scale. The peer review panels know that their scores will be translated into departmental funding and so their judgments can be considered funding allocation decisions. As stated above, given what we know about the distribution of research performance, the distribution of research output, such as papers, across UK university departments must follow a power law. Since the RAE scores assess research performance, they should also be distributed across departments in a power law. They are not.

In 2001 the RAE panels turned in a hill shaped distribution of grades to the government.  Figure 2a reports the grades assigned to all UK departments submitting to the 2001 RAE exercise, 2,598 in all. The Gini coefficient of this distribution is 0.157.[4] This set of grades was used through 2007 in allocating research resources. That the distribution of RAE grades does not reflect the known distribution of research performance; that the RAE scores take a hill shape not a power law shape and that the hill shape contains less inequality than would be expected from differences in research performance exemplifies the questions we are raising here.

The curious aspect of this exercise is the distribution of research resources that results. The UK government uses a formula to translate the grades into resource allocations. In addition to RAE grades, the formula takes into account size of department, scientific field, location in or out of London and other factors. As we are exploring the part of funding seeking to encourage research excellence through peer review judgments (and not the part focused on distributional equity) we display a stylized RAE-based resource distribution derived only from the weightings allocated to each grade. Table 1 displays the "QR weightings" used to translate grades into resources for several years. Although the grades stayed the same, the weightings changed each year. Figure 2b displays the resource distribution that would have resulted in 2002 if all departments were of the same size, in the same field, outside of London etc.  The distribution is a hill with a Gini coefficient of 0.436. The most dramatic change in QR weightings came in 2003 when the funding distribution became more reminiscent of a power law (minimum is mode, monotonic decrease to largest value),  see Figure 2c. Thereafter, minor changes were made in the weightings that accentuated inequality.[5] The Gini coefficient of the resource distribution in 2003 was 0.510 and in 2007 was 0.528+.[6]

[4] Gini coefficient calculated by translating grades onto a 1 through 7 scale.

[5] Between 2003 and 2007, the weighting given to grade 4 did not change. Grade 5 gained 14% and Grade 5* gained 30%. In addition, in 2003 a pool of money was made available to departments that had attained 5* in both the 2001 and 1996 RAE exercises. This money further accentuates the slide and is referred to as the "super 5* premium" in the table. The exact effect is not known here but is suggested by an arrow and a question mark in the graph.

[6] The empirical, non-stylized story is more complex. During this period the inequality in the funding distributed decreased because other elements in the final distribution formula were unrelated to research performance. In 2007, HEFCE commented that the increase in inequality seen here was meant to

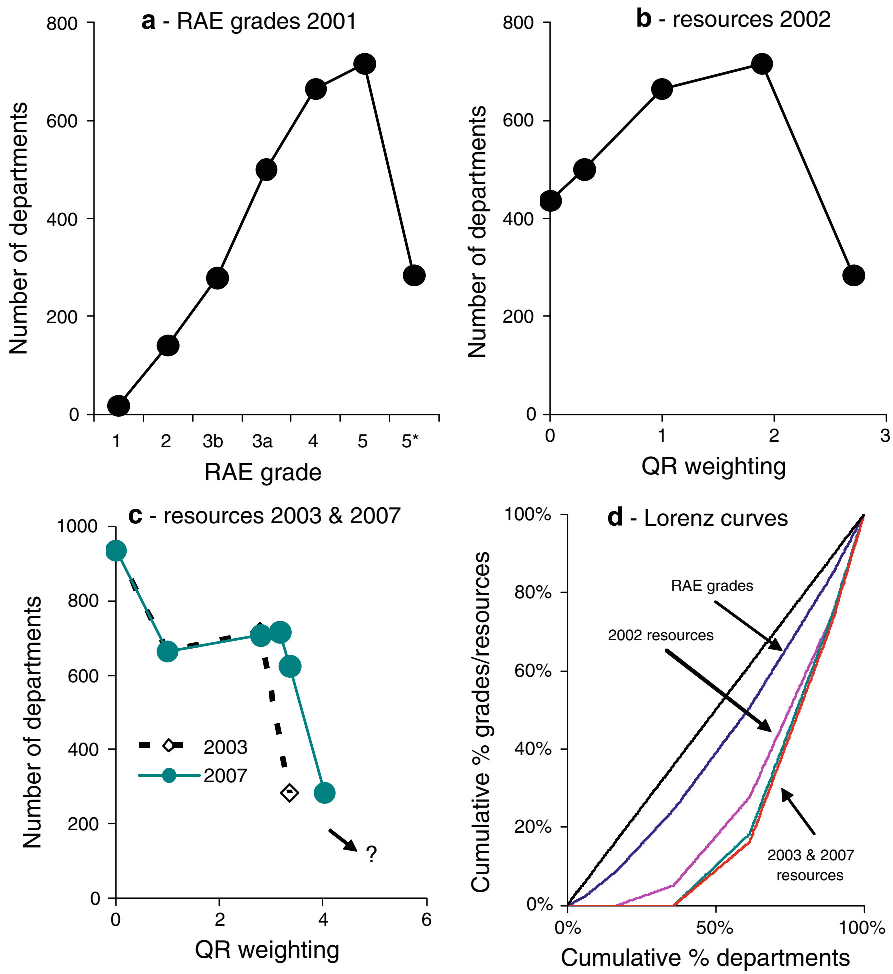


Fig. 2 UK RAE and resource distribution RAE grades obtained from HEFCE, 2007b. 2002 weights obtained from HEFCE, 2002, p. 18. 2003 weights obtained from HEFCE, 2003, p. 16. 2007 weights obtained from HEFCE, 2007a, p. 21

The RAE data support the suggestion that people making peer review judgments of research excellence may not be able to allocate resources in a way that mirrors the distribution of research performance.

Footnote 6 continued
counterbalance these other elements in the formula that were not quality weighted: "There has been a modest increase in the steepness of the selectivity scale within mainstream QR, to counterbalance the fact that RDP supervision and QR charity support funds are not quality weighted, other than for the threshold criterion that activity must be in departments rated 4 or above." From letter to Vice-chancellor Exeter from HEFCE in 2007, admin.exeter.ac.uk/ppr/GrantLetterMarch07_0119.doc.

Table 1 QR weightings used with 2001 RAE grades

Grade	2002	2003	2007
1	0	0	0
2	0	0	0
3b	0	0	0
3a	.31	0	0
4	1	1	1
5	1.89	2.79	3.175
5*	2.71	3.36	4.036 + super 5* premium

The Implications of Mismatched Distributions of Performance and Resources

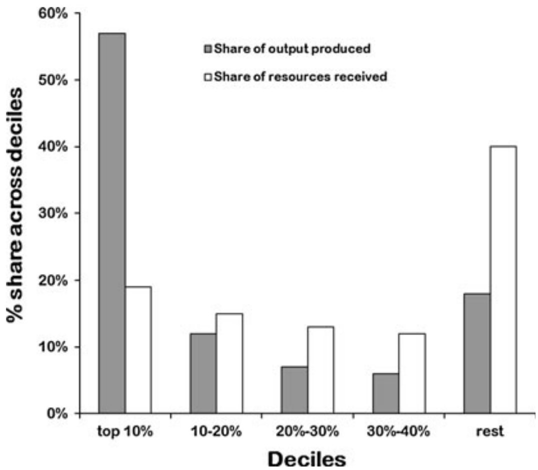
We have argued that for reasons of political values as well as our in-built aversion to inequality, there is likely to be a mismatch between the distributional form of research output and merit-based research funding allocations. Data on RAE judgments though not definitive, is publicly available and provides some support to our contention. If the argument is correct, we need to ask: so what? Would any mismatch have consequences? In this section, we examine the possibility that a mismatch in performance and resource distribution would reduce the overall research excellence obtainable in the system.

In order to examine the consequences of a mismatch, we need to model the distribution of resources. The RAE data suggest that a normal distribution would serve as an illustration. Giving everybody the same amount of research funding, which is clearly egalitarian, would violate merit-based principles of research resource distribution. To move away from egalitarian distribution, one awards better performers somewhat more money and poor performers somewhat less. The likely outcome is some variant of a Gaussian distribution.

As mentioned above, the power law distribution has a much higher degree of inequality than the normal distribution. One further point about inequality is worth considering here. Decision makers influenced by inequality aversion will be highly sensitive to the tails of the distributions they create (Amiel and Cowell 1999; Engelmann and Strobel 2004; Dawes et al. 2007). In the normal distribution above, 30 people receive \$1,000, 106 people get \$7,000 (the mean), and at the maximum one person gets \$19,000. In the power law version, 427 people get \$1,000, more than half get less than \$3,000, and at the maximum one person gets \$80,000. The minimum-maximum range of the normal distribution is one-quarter that of the power law, and the number of people stuck at the minimum in the power law is 14 times that of the normal distribution. Therefore, the normal will be preferred over the power law by inequality averse decision makers sensitized to the extremes of the distribution.

We explore the implications of the mismatch between a normal distribution of resources and a power law distribution of performance using quantiles. For illustrative purposes, power law and normal functions were constructed and divided into deciles. The power law was constructed to reflect Lotka's original formulation that 60% of the people publish one paper only. The normal distribution has mean 50

Fig. 3 Comparison of resource and output distributions



with standard deviations 30. For the top four deciles and the rest, Figure 3 reports the share of output produced assuming a power law distribution and the share of resources received assuming a normal distribution

First, note that in a true egalitarian distribution of funding, the top 40% of researchers would receive 40% of research resources and the bottom 60% would receive 60% of resources. In contrast, the normal distribution of resources gives the top 40% most productive researchers more funding (59%) and the bottom 60% of researchers less funding (41%) than they would receive under an egalitarian distribution in which every decile would receive 10% of funding. The normal distribution is thus a merit-based distribution of funding.

Although the world in which productivity followed the Lotka distribution and research resources were distributed normally might improve upon the world of flat distribution of funding, we can still find reason to question the resource distribution. In this scenario, the top 10% of research performers would split 20% of the available research resources, which is more than any other decile, but far less than the 60% they might expect if resources were commensurate with their output. In the second decile in contrast, researchers split 15% of the resources, which is more than the 12% of output they produce. The same is true for the rest of the deciles in the top 40%; in each decile the researchers split a greater share of resources than they produce in output.

To the extent that obtaining more resources for research[7] provides any motivation and incentive for researchers in their work, researchers might want to avoid the top 10% because the rewards are so much less than the output produced. But we can say more. Assume that attaining the top 10% is more trouble than it is worth, where might strategic researchers aim to position themselves? The best place appears to be the 10-20% decile because researchers in that decile split more resources than any below them, while still earning a premium on their output share. In other words, the

[7] Not salary increases, but rather grant funding to buy equipment and hire students and post-docs and pay summer salary.

proposition here is that researchers might aim high but not too high in a system where resources are distributed by inequality averse decision makers.

This proposition finds some qualitative support in real life situations. Gläser et al. reviewed responses to the UK RAE system and argued that the literature can be read as suggesting that the RAE (and indeed other evaluation-based methods of research funding) "improve quality to the upper middle level and drive out low quality research but suppress excellence to a certain extent" (Gläser et al. 2002, p. 22). This analysis confirms the suspicion of RAE observers in that the 10-20% decile is indeed "upper middle" and appears to be a more desirable location than achieving excellence among the top 10% in a system in which resources are awarded on merit-based judgments which do not fully mirror the distribution of research performance.

The concern of the National Science Foundation with transformative research arguably expresses the same worry. Arden Bement, former director of NSF, explained transformative research as "a range of endeavors, which promise extraordinary outcomes; such as, revolutionizing entire disciplines, creating entirely new fields, or disrupting accepted theories and perspectives" (Bement 2007).[8] If the output distribution discussed here is seen as a distribution of impact, transformative research can be interpreted as directing attention to the very high end of the long tail in the power law output distribution. Implicit in current NSF initiatives to enhance their support of transformative research is an assessment that they were previously failing to support enough transformative research. This can be restated as: NSF is worried that their evaluation-based methods of research funding suppress excellence to a certain extent.

Discussion

In this paper, we propose that there is a tension between distributions of research performance and merit-based research funding decision making. In effect we argue that decision makers who increase concentration of resources because they accept that research performance differs, and that resources should be distributed according to merit, probably implement less inequality than might be justified by differences in research performance. If they did they would be accused of fostering "undue concentration." Any resource distribution that truly mirrored the distribution of research performance would be untenable in the public realm because of its extreme degree of inequality.

Our argument is stylized and there are several factors that may mitigate this effect in real life. The first is the observation that after producing a particularly noteworthy breakthrough, researchers, at least in the U.S., can find themselves in the position of having more money than they can productively spend. In other words, real life laboratories can become saturated with resources that arrive after the fact.[9]

[8] Bement also expects the pursuit of transformative research to be high risk. There is overlap in the formal analysis of risk and inequality, therefore it is not a surprise that when people discuss something distributed so unequally as is research output and impact, that they regularly note the unaccustomed level of risk involved.

[9] Personal communication, Juan Rogers, January 2008, based on interviews with researchers responsible for key discoveries in nanotechnology.

That the saturation level of resources is less than commensurate with the laboratory contribution to the progress of knowledge may make this argument moot. Others would disagree, for example John Reed, the most highly cited scientist in cell biology and current holder of 11 NIH grants worth almost $11 million, who states that: “The evidence is that some labs and some people can handle a larger portfolio” (quoted in Hand 2008).

The second factor that runs counter to the proposition is that real life researchers may seek status rather than monetary return to their endeavors. Status, as expressed in prizes, invitations to speak, job offers and general acclaim (if we could measure it accurately) might well follow a power law distribution as it does not seem like it would be subject to the pressures of inequality aversion and distributional equity that attach to material resources. If real life researchers are concerned only with status, and work in a system with many visible markers of status, the details of monetary returns to performance may not be particularly salient, given that in every merit-based system they see some monetary return to their high performance. However, prizes and acclaim alone do not enable a person to conduct further research. Those motivated to continue in research and indeed to compete at the highest level in their fields need the time, equipment and talent that funding buys.

Conclusions

In this paper, we have examined a fundamental tension in science policy, that between equity and excellence. Although using merit-based evaluation as a criterion for research funding would seem to resolve this tension, we argue here that this is unlikely. Merit-based decision making alone is insufficient because of inequality aversion, a fundamental tendency of people to avoid extremely unequal distributions. The distribution of performance in science is extremely unequal, and a decision maker distributing public money in a way that matched the level of inequality in performance would be accused of undue concentration. In fact, decision makers are likely unaware of the issue, as they no doubt operate with distributional assumptions of normality that guide our everyday intuitions.

This argument may matter to those who seek to optimize both the research excellence in a system and efficiency of funding allocations. If all researchers were indifferent to the amount of money they received for their research, or if rewarding research performance with more resources to conduct research destroyed everybody’s intrinsic motivation and so reduced motivation, or if it were impossible to distinguish the top 10% of researchers from the next decile, these quantitative relationships would have no real world import. However, if even some researchers seek to maximize the resources available for their research activities, and expect allocation of resources to relate to their achievements and effort, then these ideas may be relevant to overall system performance.

We did not argue that an egalitarian distribution of resources would produce an even distribution of output, any distribution of resources would result in a power law distribution of output. However, if one wants to optimize system performance with scarce resources, identifying the best resource distribution becomes relevant.

Because research output is a power law and this is a fat tailed distribution, overall system performance will be highly dependent on the top of the distribution. This is precisely where the mismatch between normal resource distribution preference and power law performance distribution is greatest because it is where people are most sensitive to resource inequality.

However, the arguments in the paper will remain conjectures until properly constructed empirical comparisons of funding and performance distributions can be analyzed over time and across countries.

Acknowledgements The authors would like to thank the many colleagues whose valuable comments helped shape this paper, including Christian Houdre, Juan Rogers, Doug Noonan, John Walsh, and participants in a seminar at University of Twente.

References

Amiel, Yoram, and Frank Cowell. 1999. Thinking about Inequality. Cambridge: Cambridge University Press.

Bement, Arden L. Jr. 2007. Transformative Research: The Artistry and Alchemy of the 21st Century, speech given at the Texas Academy of Medicine, Engineering and Science Fourth Annual Conference, Austin, Texas, January 4, 2007, available at: http://www.nsf.gov/news/speeches/bement/07/alb070104_texas.jsp, downloaded on January 11, 2008.

Clauset, Aaron, Cosma Rohilla Shalizi and M.E.J. Newman. 2009. Power-Law Distributions in Empirical Data. *SIAM Review* 51: 661–703. (Preprint at arxiv:0706.1062).

Dawes, Christopher T., James H. Fowler, Tim Johnson, Richard McElreath and Oleg Smirnov. 2007. Egalitarian motives in humans. *Nature* 446: 794-796.

Dorfman, Robert. 1979. A Formula for the Gini Coefficient. *The Review of Economics and Statistics* 61: 146-149.

Engelmann, Dirk, and Martin Strobel. 2004. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *The American Economic Review* 94: 857-869.

Feller, Irwin. 2001. Elite and/or Distributed Science. Chapter 9 in Maryann P. Feldman and Albert N. Link (eds.) *Innovation Policy in the Knowledge-Based Economy*. Dordrecht, The Netherlands: Kluwer Academic.

Gläser, Jochen, Grit Laudel, Sybille Hinze and Linda Butler. 2002. Impact of evaluation-based funding on the production of scientific knowledge: What to worry about and how to find out. Expertise for the German Ministry for Education and Research, May, http://repp.anu.edu.au/expertise-glae-lau-hin-but.pdf.

Geuna, Aldo. 1998. Resource Allocation and Knowledge Production: Studies in the Economics of University Research. Dissertation no. 98-41, Faculty of economics and Business Administration Universiteit Maastricht, The Netherlands.

Hand, Eric. 2008. Day in the Life of an 11 Grant Grandee. *Nature* 452: 259.

HEFCE 2002. Funding higher education in England: How the HEFCE allocates its funds. HEFCE 02/18, http://www.hefce.ac.uk/pubs/hefce/2002/02_18/02_18contents.pdf.

HEFCE 2003. Funding higher education in England: How the HEFCE allocates its funds. HEFCE 03/29, http://www.hefce.ac.uk/pubs/hefce/2003/03_29/03_29.pdf.

HEFCE 2007a. Funding higher education in England: How the HEFCE allocates its funds. HEFCE 07/20, http://www.hefce.ac.uk/Pubs/hefce/2007/07_20/07_20.pdf.

HEFCE 2007b. Funding data for quality-related research QR. funding: 2007-08 Quality-related research funding. http://www.hefce.ac.uk/research/funding/qrfunding/, data as at March 1, 2007.

Lotka, Alfred J. 1926. The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences* 16: 317-323.

Morin, Alexander J. 1993. *Science Policy and Politics*. Upper Saddle River, NJ: Prentice-Hall.

Tricomi, Elizabeth, Antonio Rangel, Colin F. Camerer and John P. O'Doherty. 2010. Neural evidence for inequality-averse social preferences. *Nature* 463: 1089-1091.

Wu, Yonghong. 2010. Tackling undue concentration of federal research funding: An empirical assessment on NSF's Experimental Program to Stimulate Competitive Research (EPSCoR), *Research Policy* 39: 835-841.